Meta-research: The effect of the COVID-19 pandemic on the gender gap in research productivity within academia

Kiran G. L. Lee1,2, Adele Mennerat3, Dieter Lukas4,§, Hannah L. Dugdale1,§, Antica Culina5,6,§,

1 Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

2 Department of Animal & Plant Sciences, University of Sheffield, Sheffield, UK

3 Department of Biological Sciences, University of Bergen, Bergen, Norway

4 Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

5 Rudjer Boskovic Institute, Zagreb, Croatia

6 Netherlands Institute of Ecology, NIOO-KNAW, Wageningen, The Netherlands

§ Shared last authorship

Corresponding Authors:

Kiran Gok Lune Lee

School of Biosciences, Alfred Denny Building, Western Bank, Sheffield, S10 2TN, UK

Email address: [kgllee1@sheffield.ac.uk](mailto:kgllee1@sheffield.ac.uk)

Antica Culina

Rudjer Boskovic Institute, Zagreb, Croatia

Email address: [aculina@irb.hr](mailto:aculina@irb.hr)

# Abstract

Academic success is a commonly evaluated by the number of published articles. Yet, this metric selects against female academics. Novel social conditions induced by the COVID-19 pandemic likely exacerbated this gender disparity if female academics took on the load of caregiving, domestic, service and teaching roles. We investigate the overall pandemic effect on the gender gap in research productivity through a systematic review and meta-analysis of 130 effect sizes from 55 published articles across scientific disciplines. We also investigate three hypotheses on how research field, breadth of gender gap before the pandemic, and authorship position influence this effect. We found the gender gap in research productivity within academia has overall increased during the COVID-19 pandemic, especially in social sciences and medicine, fields that were previously nearest to being gender equal. We did not detect an influence of authorship position on the effect. Our results further corroborate calls to improve the support and evaluation systems in academia.

Subjects: Academia, Gender Disparity, Research Productivity

Key words: COVID-19, Pandemic, Academia, Gender disparity, Women researchers, Research productivity

# Introduction

Research productivity, defined as the number of manuscripts or publications, is a traditionally used, but flawed metric for evaluating academic merit because it biases against individuals according to socio-demographic circumstances. Women are disadvantaged compared to men when success is measured using traditional metrics of research productivity, despite no actual differences in contribution and impact of research (van den Besselaar and Sandström, 2016, 2017; Astegiano, Sebastián-González and Castanho, 2019; Huang *et al.*, 2020). Additionally, during the COVID-19 pandemic, novel living and working conditions worsened the research productivity, of many women worldwide (Anwer, 2020; Boncori, 2020; Guy and Arthur, 2020; Herman *et al.*, 2021; Altan-Olcay and Bergeron, 2022).

Multiple factors are likely to contribute to gendered changes to research productivity during a pandemic. First, women generally perform more unpaid caregiving and domestic work (Schiebinger, Henderson and Gilmartin, 2008; Schiebinger and Gilmartin, 2010). Social-distancing and facility closures during the pandemic increased caregiving and domestic work (Carli, 2020; Carlson, Petts and Pepin, 2020) with reduced community help from nurseries, schools, care homes, house cleaners, laundrettes, nannies, babysitters and family (Myers *et al.*, 2020; Barber *et al.*, 2021; Breuning *et al.*, 2021; Deryugina, Shurchkov and Stearns, 2021; Shalaby, Allam and Buttorff, 2021). As these tasks have disproportionately fallen on women, time and space for academic research during “work-from-home” conditions was difficult (Abdellatif and Gatto, 2020; Boncori, 2020; Guy and Arthur, 2020). Second, the distribution of work within academic institutions is often gendered. Women undertake more ‘non-promotable’ tasks (Babcock *et al.*, 2022) such as administrative, supportive and mentoring roles (Porter, 2007; Mitchell and Hesli, 2013; Babcock *et al.*, 2017; Guarino and Borden, 2017; O’Meara *et al.*, 2017; O’Meara, Kuvaeva and Nyunt, 2017). Changes in teaching and administration in response to the pandemic were therefore more likely to be facilitated by women (Docka-Filipek and Stone, 2021; Minello, Martucci and Manzo, 2021). Third, labour roles contributing towards publication are also gendered. Women generally perform more technical work such as generating data, whilst men assume more core tasks in conceptualisation, analysis, writing and publishing (Macaluso *et al.*, 2016). Pandemic closures to research institutions would therefore likely impact women authorship from technical roles stronger than men. Additionally, the surge in publications during the pandemic (Else, 2020) could have reduced the quality of peer review, with evaluation being more influenced by cognitive shortcuts. These shortcuts are often associated with biases tending to operate against women (Kaatz, Gutierrez and Carnes, 2014; Reuben, Sapienza and Zingales, 2014; Carli, 2020) resulting in lower success getting submissions accepted (Fox and Paine, 2019; Murray *et al.*, 2019; Day, Corbett and Boyle, 2020; Hagan *et al.*, 2020).

The role of these factors shaping the gender gap in research productivity during the pandemic might differ across research fields (Madsen *et al.*, 2022). One possibility is that research fields that were already more gender-disparate may have experienced the most exacerbated gender gaps during the pandemic. In fields that were already traditionally more gender-disparate, less support may have been available to women to balance the effects of the pandemic. Male-dominated fields often lack viewpoints of female colleagues, and might therefore be less likely to identify and support paid care work or extended leave options (Clark, 2020; Nash and Churchill, 2020). An alternative possibility is that the pandemic might have eroded the support structures that existed in more gender balanced fields. The pandemic may also have exacerbated a gender gap in authorship position (first, middle or last) (King and Frederickson, 2021) if additional service, teaching, caregiving, and domestic roles taken up by female academics during the pandemic may limit their abilities to perform research (as first authors) or lead research (as last authors) but not in supporting research (as middle authors).

Here, we quantitatively calculated by meta-analysis the overall effect of the COVID-19 pandemic on the gender gap in research productivity and predicted the gap increased compared to the period just prior to the pandemic, such that male academic productivity saw even further increases. As studies differ in the type of research productivity measured, between individual survey responses, numbers of submissions and numbers of publications, we investigated the influence this might have on the gender gap increase observed during the pandemic, but with no expectation of any differences. Second, we explored variation in the gender gap increase across research fields and then explored the effect of research field according to the previous degree of gender disparity. We predicted the gender gap is exacerbated in fields that already had a previously greater gender gap, as according to the proportion of female authors, because of less support available to women to balance the effects of the pandemic. Third, we explored whether the disparity in favourable authorship positions has increased. We predicted the gender gap has increased more in first and last, rather than middle authorship positions because female academics have been especially more limited in undertaking leading, but not supportive research roles in lockdown conditions.

# Results

Our systematic literature review identified 55 studies that met the inclusion criteria (for details on the procedure please see the Methods section). All of the studies we were able to find only compared women to men (see Limitations). We extracted and calculated 130 effect sizes from these studies and performed a meta-analysis and meta-regression to test our three hypotheses and related predictions.

### 1a: Has the pandemic increased the gender gap in research productivity across academia?

Across all samples, after controlling for multiple effect sizes from the same study, we found the relative productivity of women to men decreased during the pandemic at -0.071 compared to before the pandemic (95% CI = -0.099 to -0.043, SE = 0.0144, p = <0.001, Fig. 1). There is large variation in the 130 effect sizes, with 38 indicating a clear increase in the gender gap (95% confidence intervals within negative ranges) and 56 a trend of an increase (Effect size is negative but 95% confidence intervals are not within negative ranges), while 11 indicate a clear decrease in the gender gap (95% confidence intervals within positive ranges) and 25 a trend of a decrease (Effect size is positive but 95% confidence intervals are not within positive ranges). Total heterogeneity was high (*I2*= 97.9%), with 46.6% of it explained by whether research productivity was measured by survey responses or submission/publication numbers and 52.1% explained by the individual effect sizes.

Chart, scatter chart

Description automatically generated

**Figure 1.** **Overall effect of the pandemic on academia’s gender gap in productivity.** Orchard plot showing all 130 effect sizes (points) and their weight (point size), with the mean effect size (darker coloured point outlined black and vertically centred), the 95% confidence interval (horizontal thick black bar) and the 95% prediction interval of the expected spread of effect sizes based on between-study variance (horizontal thin black bar). k =x(y), where x is the number of effect sizes and y is the number of studies.

### 1b: Does the type of research productivity measure influence how much the pandemic has changed the gender gap in research productivity?

The degree of increase in the gender gap caused by the pandemic differed according to the type of research productivity measured (QM(df = 3) = 37.130, p < 0.001, Fig. 2). Studies measuring changes to research productivity during the pandemic based on surveys detected a larger overall effect (-0.192, 95% CI = -0.272 to -0.113, SE = 0.041, p < 0.001) than studies that compared the number of articles published (-0.047, 95% CI = -0.085 to -0.008, p = 0.017, SE = 0.020) or submitted (-0.053, 95% CI = -0.087 to -0.018, p = 0.003, SE = 0.017) by authors of each gender before and during the pandemic.

A picture containing text, screenshot, diagram

Description automatically generated

**Figure 2. Type of research productivity measure influence on academia’s gender gap during the pandemic.** Orchard plots comparing the distribution of effect sizes (points) and their weight (point size) depending on the type of research productivity measured for which the mean effect size (darker coloured points outlined black and vertically centred), the 95% confidence interval (horizontal thick black bar) and the 95% prediction interval of the expected spread of effect sizes based on between-study variance (horizontal thin black bar) is given. k= x(y), where x is the number of effect sizes and y is the number of studies.

### 2a: Has the pandemic affected women differently across research fields?

We found little evidence of a significant differential impact of research fields on the reported effect sizes (QM (df = 5)= 21.967, p = 0.001, Fig. 3). When considering research fields individually, social sciences showed the greatest increases in the academic productivity gender gap during the pandemic (-0.084, 95% CI = -0.143 to -0.024, SE = 0.030, p = 0.006), followed by medicine (-0.066, 95% CI = -0.102 to -0.029, SE = 0.019, p < 0.001). The pandemic showed little effect in multidisciplinary fields (-0.050, 95% CI = -0.125 to 0.024, SE = 0.038, p = 0.188), biological sciences (-0.003, 95% CI = -0.057 to 0.050, SE = 0.027, p = 0.902) or Technology, Engineering, Mathematics, Chemistry and Physics fields (-0.003, 95% CI = -0.053 to 0.048, SE = 0.026, p = 0.916).

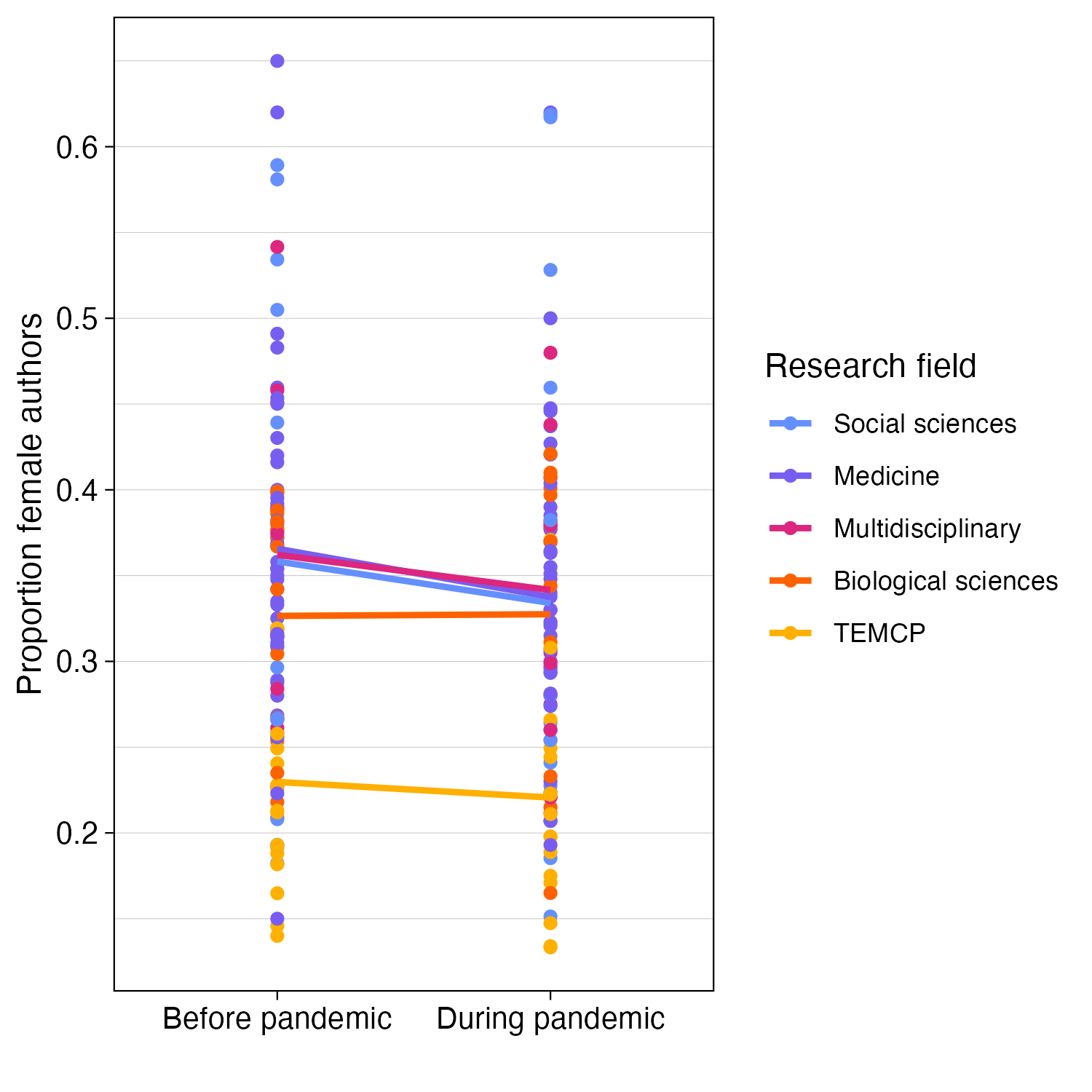
A picture containing text, screenshot, map, diagram

Description automatically generated

**Figure 3.** **Research field influence on academia’s gender gap during the pandemic.** Orchard plot comparing the distribution of effect sizes (points) and their weights (point sizes) depending on the research fields sampled, for which the mean effect size (darker coloured points outlined black and vertically centred), the 95% confidence interval (horizontal thick black bar) and the 95% prediction interval of the expected spread of effect sizes based on between-study variance (horizontal thin black bar) is given. TEMCP stands for Technology, Engineering, Mathematics, Chemistry and Physics. k =x(y), where x is the number of effect sizes and y is the number of studies.

### 2b: Has the pandemic exacerbated existing differences in gender disparity across research fields?

The pandemic has increased the gender gap in article output more in journals/repositories/pre-print servers that were previously less gender-disparate (QM(df = 1) = 10.285, p = 0.001). When grouping studies by research fields (Fig. 4), those with a smaller gender gap prior to the pandemic experienced greater increases in the gender gap in academic productivity during the pandemic compared with fields where the gender gap was already large to start with (Social sciences: 35.8% to 33.4%, medicine: 36.6% to 33.8%, multidisciplinary: 36.2% to 34.2%, biological sciences: 32.7% to 32.7%, Technology, Engineering, Mathematics, Chemistry and Physics: 23.0% to 22.1%).



**Figure 4. Research field influence on academia’s gender gap in submissions and publications during the pandemic.** Line plot of gender gap in authorship before and during the pandemic, grouped by research field. Points show gender gap as the proportion of female authors publishing and submitting before or during the pandemic and are coloured according to research field. Lines take the mean value of these points according to research field. TEMCP stands for Technology, Engineering, Mathematics, Chemistry and Physics.

### 3a: Has the pandemic affected women more in their ability to lead rather than support research?

We found no evidence of a significant differential impact of authorship position on effect sizes (QM(df = 5) = 13.190, p = 0.022, Fig. 5).

A picture containing text, screenshot, diagram, plot

Description automatically generated

**Figure 5.** **Authorship position and academia’s gender gap during the pandemic.** Orchard plot comparing the distribution of effect sizes (points) and their weight (point sizes), depending on the authorship position sampled in publication studies, for which the mean effect size (darker coloured points outlined black and vertically centred), the 95% confidence interval (horizontal thick black bar) and the 95% prediction interval of the expected spread of effect sizes based on between-study variance (horizontal thin black bar) is given. k= x(y), where x is the number of effect sizes and y is the number of studies.

### Is there evidence of publication bias?

The multilevel meta-regression, including article-output studies and survey studies as a moderator because of their differences in sample size, showed no evidence of publication bias (Article-output studies: slope= -0.025, 95% CI = [-0.059- 0.009], SE = 0.017,  p = 0.148; Survey-response studies: slope= -0.157, 95% CI = [-0.244-  -0.071], SE = 0.044,  p <0.001). This model correlates standard error with effect size and negative slopes suggest small studies do not have large effect sizes. meta-regression. A visual inspection of the funnel plots (Fig. 6) similarly did not indicate any suggestion of publication bias.

### Are our results robust?

Our results show little change in a sensitivity analysis that excludes seven effect sizes using four measures of productivity from survey-based studies that are less directly comparable: research time (N = 4), job-loss (N = 1), burnout (N = 1) and the number of projects (N = 1). When excluding these effect sizes, the overall estimate in prediction 1a (overall pandemic effect on gender gap) is -0.063, 95% CI = [-0.0892- -0.0372] , SE = 0.0133, p < 0.001 and the estimate for survey-studies in prediction 1b (method of research productivity influence on pandemic effect on gender gap) is -0.239, 95% CI = [-0.3419- -0.1352], SE = 0.0527, p < 0.001.

The model investigating prediction 1a (the overall effect of the pandemic) still showed a significant effect in a leave-one-out analysis, which repeatedly fitted the model leaving out one effect size at a time (Supplementary file 3). Leaving out the most influential effect size, the overall estimate is -0.058, 95% CI = [-0.081 - -0.035] , SE = 0.012, p < 0.001.

Diagram

Description automatically generated

**Figure 6. Investigating publication bias.** Funnel plot of effect sizes and their precision, as a function of standard error, from studies that measure research productivity by (a) survey-responses (b) article output. The vertical dashed line is the summary effect size. The legend outlines levels of statistical significance for effect sizes based on their precision.

# Discussion

Our study finds quantitative evidence, based on 55 studies and 130 effect sizes, to support the hypothesis that the COVID-19 pandemic has exacerbated gender gaps in academic productivity. These findings are consistent with the notion that novel social conditions induced by the pandemic have disadvantaged women in academia even more than before. Overall, the studies summarised in our meta-analysis suggest that the gender gap in research productivity within academia has increased during the pandemic by 7%. We found no evidence of a publication bias. There is high heterogeneity in the effect sizes reported from different studies, arising from the type of research productivity measured. When measuring research productivity as the number of published or submitted articles, we find a slightly smaller increase in the gap of around 5%. This corresponds to the proportion of authors on submitted or published articles who are women declining from an average of 33.2% pre-pandemic to 31.4% during the pandemic (-0.05\*33.2%=-1.7%). Such a change might reflect lower submission and acceptance rates of articles by women compared to their male colleagues or an increased drop-out of woman from academia caused by the pandemic. When measured by survey’s women in academia reported the pandemic reduced productivity 17% more than for male counterparts. Future studies might detect different effects of the pandemic on gender disparity in productivity. As our data and analytical codes are open, and the literature on pandemic effects is increasing, we hope our work can form a first step in a living systematic review on the topic.

Our study likely underestimates the pandemic effect on article productivity in women because writing and publishing can take a long time (Powell, 2016). Many of the articles submitted or published during the pandemic were likely started and at least partially completed prior to the pandemic, given that most research grants span multiple years. With restricted access to laboratories, field sites and collaborators, many new projects have been delayed (Corbera *et al.*, 2020), making it likely that the article-output studies we could include by the time our study started in 2021 underestimates the true effect of the pandemic, which might span over many years. In support of this view, we find some indication for a larger, real-time effect from the effect sizes based on survey responses, which indicate a much stronger negative effect of the pandemic on women’s productivity compared to men’s (effect size = -0.192). This signals that women are one fifth more likely than men to indicate that the pandemic has negatively affected their academic activities, which may stem from a combination of women on average feeling a larger strain, and a larger proportion of women being severely affected by the pandemic. In the literature used within our meta-analysis, five of six survey studies report evidence of a negative interaction effect of being both female and a parent on research productivity during the pandemic, presumably because of increased caregiving demands. Effect sizes are highly varied in survey response studies, which may reflect subtle differences in the measure of research productivity asked in the survey or some populations of survey respondents holding strong opinions of the pandemic As with all survey studies, their data can be subjective or without evidence and suffer from limited numbers of respondents sampled which may inflate the overall estimate

Our analysis suggests the pandemic may have differentially impacted female researchers across research fields, with increases in gender gaps particularly visible in research fields that were nearest gender-equality before the pandemic. Social sciences and medicine were two fields closest to gender equality that experienced the most significant decrease in female authors. Female researchers working in fields with previously gender-equitable environments may have experienced new, difficult research conditions induced by the pandemic, whereas in gender-biased fields, these difficulties might already have been present. For example, financial structures or childcare arrangements that could previously assist parents with caregiving responsibilities may have broken down during the pandemic (Fortin and Taylor, 2020; National Academies of Sciences, 2021). Alternatively, social sciences and medicine are fields that could have had the greatest surge in COVID-19 and pandemic-related research. Women in social sciences and medicine potentially had less opportunities to pursue this new pandemic-related research because of extra work performed in gender roles, or because women already had relatively smaller collaborative networks, fewer senior positions and less funding. Additionally, many medical journals sped up the publication process (Horbach, 2020), so the real-time effect of the pandemic on research productivity in women versus men may be reflected more in papers submitted and published in medicine than in other fields.

We did not find a clear signal that disparity in research productivity differed according to authorship positions on submitted/published articles. We were limited in addressing this prediction because the samples used for calculating effect sizes for particular authorship positions were too small to infer whether there was a differential change in the gender gap. Additionally, not all fields have the same authorship order norms making comparisons difficult to study.

It seems unlikely that this increase simply represents a normal temporal fluctuation. The survey results, which report the strongest effects, specifically focused on the influence of the pandemic above and beyond the pressures researchers might already normally experience. The 5% decline in the proportion of authors who are women also likely indicates the extraordinary circumstances of the pandemic. This decline is steep relative to a study comparing the change in the proportion of female authors between 1945 and 2005 showed a steady increase from 14% of all authors being women to 35%, with no apparent year-on-year decline since at least 1990 (Huang et al. 2020).

Our study has several limitations, which are outlined in the Limitations of the methods and summarised here. Causes of the increased gender disparity in research productivity during the pandemic are not identified. Gender is investigated only as a binary variable, mostly using first-name prediction tools, with no investigation into non-binary or transgender. Geographic regions are not investigated, limiting generalisability as our samples are not equally representative of all geographic regions. All studies are to some extent unique by combined differences in sampling method and analyses performed.

# Conclusion

Overall, our study highlights an exacerbated gender gap in academic research productivity during the COVID-19 pandemic. This gender gap was exacerbated more in social sciences and medicine, which are fields that were previously less gender-disparate and may represent regression in progress made towards gender equality. Academic institutions should acknowledge and carefully accommodate the pandemic period when using research productivity to evaluate female academics for career progression in the coming years. For example, tenure-clock extensions designed to accommodate pandemic disruption may inadvertently exacerbate the gender gap by extending the period that advantaged individuals can outperform. More emphasis must be placed evaluating academic merit using more holistic measures and on an individual basis. We recommend future studies to investigate potential mechanisms of the increased gender gap in academic productivity, to continue monitoring the gendered disparity of academic productivity, and to investigate any long-term implications that can arise from reduced productivity of women.

# Methods

### Search process

We carried out a systematic review to identify, select and critically evaluate relevant research through data collection and analysis. We reported it following PRISMA guidelines (Moher *et al.*, 2009). We carried out the literature search process in three steps: 1) a scoping search, 2) an initial search with pre-selected author terms, and 3) a refined search using terms as recommended by the litsearchR 1.0.0 (Grames *et al.*, 2019). All searches were filtered for texts from 2020 onwards. We initially performed a scoping search to determine if there were over ten texts with primary research investigating differences by gender in academic productivity before and during the pandemic. The scoping search was conducted on 30/06/2021 by Google searching combinations of synonyms for: 1) the COVID-19 pandemic, 2) gender, 3) academia, 4) inequality and 5) productivity. The scoping search identified 21 original research publications with quantitative metrics investigating differences in academic productivity by gender before and during the pandemic (scoped texts). Of these 21 studies, 14 were indexed by Web of Science, and 17 (including the same 14 from Web of Science) were indexed by Scopus.

Terms for the initial search were selected by scanning the title, abstract and keywords of scoped texts. We constructed an initial Boolean search string according to the PICO (Population, Intervention, Comparator, Outcome) framework (Livoreil *et al.*, 2017). Population was represented by “academia”, Intervention by “pandemic”, Comparator by “gender” and Outcome by “inequality” and “productivity” (Table S1). A sixth concept group contained terms used to exclude irrelevant studies that did not investigate studies in hypothesis one. Terms within concept groups were connected by the Boolean OR operator, and the concept groups were connected by the AND or AND NOT operators, enabling searches for any combination that includes one term from each of the six concept groups. Terms in the initial search were selected by scanning the title, abstract and keywords of scoped texts. The initial search in Scopus generated 722 texts, including 13/17 (76.5%) of scoped texts indexed by Scopus.

To improve the 13/722 (1.8%) efficiency of finding scoped texts from our initial search, we imported all 722 texts into R and used litsearchR. Using litsearchR, potential key terms were extracted from the title, abstract and keywords of texts using the Rapid Automatic Keyword Extraction algorithm. A ranked list of important terms was then created from building a key term co-occurrence network (Table S2). Six high-strength terms within the key term co-occurrence matrix, describing research not relevant to our study, such as those of an epidemiological or experimental nature, were added to the AND NOT operator concept group to exclude texts mentioning these terms. Table 1 describes terms of the refined Boolean search string and their respective concept groups. We performed the refined search on 27/07/2021 and generated 700 total texts combined from Scopus (126 texts, including 13/17 studies found in the scoping search), the Web of Science core collection (199 texts), EBSCO (276 texts and Proquest (99 texts) from 2020 onwards. The final search hit rate had an efficiency of 10.3% (13/126) on Scopus. After removing duplicates, 580 texts remained to enter the study screening stage.

### Study screening

To be included in our meta-analysis, studies had to quantitatively investigate gender differences in productivity within academia before and during the pandemic. In our initial screen of titles, abstracts and keywords of studies we kept only those suggesting the study investigated: 1) academia, 2) gender, 3) the COVID-19 pandemic and 4) some measure of productivity (Table S3). We included any text returned by the database, including grey literature in any language, though all texts had abstracts with an English version. To ensure repeatability of the screening process, we used Rayyan.ai (Ouzzani *et al.*, 2016) to blind the inclusion or exclusion of 420 randomly selected studies by two reviewers (K.L. and D.L.). The agreement rate between reviewers was 97%, with 49 studies that both authors agreed to include, 357 studies which both excluded, ten studies one reviewer included but the other excluded, and 4 studies only included by the other reviewer. This agreement rate resulted in a, “strong” (McHugh, 2012) to “near perfect” (Landis and Koch, 1977) Cohen’s kappa of 0.86. Of the 14 studies which were included by one but excluded the other, 3 were included after joint review. To reduce workload, we did not double-initial-screen 160 texts because of 97% agreement rates achieved, and of these K.L. included 18 and excluded 142 articles. Overall, out of the 580 texts, 70 were retained (Fig. S1) for the full text screening.

Full texts were single-screened by K.L. by following hyperlinks to studies and recording in a spreadsheet studies had the following qualifiers: 1) for both genders, 2) some quantifiable metric of academic productivity measured, and 3) for before the pandemic compared with during the pandemic (time periods chosen according to the authors’ discretion) , 4) using primary research, 5) with full texts available. Studies were not assigned quality weightings. Thus, 25 studies that all contained necessary metrics to calculate effect sizes were retained for data extraction, excluding 45 studies (Fig. S1).

### Iterating the search

To find studies that had been published since the 27/07/21 search (Table 1), we iterated the search and screen process. The second search was repeated on 28/02/2022, generating 1646 total texts combined from Scopus (258 texts, including 14/17 studies found in the scoping search), the Web of Science core collection (413 texts), EBSCO (542 texts) and Proquest (433 texts) from 2020 onwards. We removed 438 duplicates using Rayyan.ai, leaving 1208 de-duplicated studies. To ensure our methods are repeatable, we checked and found all 580 de-duplicated studies from the previous search were also found again. Out of the 1208 texts from the final search, we included 170 after screening titles, abstracts and keywords. For these 169, we screened the full texts, excluding 120 studies (53 studies missed one qualifier, 38 studies missed two qualifiers, 27 studies miss three qualifiers, 1 study missed four qualifiers) and keeping 50 (including the 25 identified in the original search) that all contained the necessary information to calculate the effect sizes (see Supplementary file 1). Two studies with full texts in Spanish were translated to English by Google translate, of which one was included. Five studies found in the scoping search but not returned in database searches were also included, resulting in 55 total studies used to extract variables for analysis.

**Table 1. Final Boolean search string used in full literature search for texts** **since 2020.** Terms in italics were added using litsearchR.

|  |  |  |
| --- | --- | --- |
| Concept group | PICO group | Terms |
| Academia | Population | ( academi\* OR author\* OR database\* OR journal\* OR research OR scien\* ) |
| Gender | Population | AND  ( female\* OR gender OR male\* OR men OR women ) |
| Pandemic | Intervention | AND  ( coronavirus OR covid OR pandemic ) |
| Inequality | Comparator | AND  ( bias\* OR disparit\* OR disproportion\* OR fewer OR gap OR "gender difference\*" OR imbalance\* OR inequalit\* OR inequit\* OR parity OR "sex difference\*" OR skew\* OR unequal ) |
| Productivity | Outcome | AND  ( performan\* OR publication\* OR publish\* OR productiv\* ) |
| Exclusion of biomedical studies | Population | AND NOT  (experiment OR laboratory OR mortality OR surviv\* OR *"acute respiratory"* OR *gis* OR *icu* OR *risk* OR *rna* OR *symptoms* ) |

### Our sample

Our sample consists of 130 effect sizes obtained from 55 studies including surveys of potentially affected people (Myers *et al.*, 2020; Rodríguez-Rivero *et al.*, 2020; Barber *et al.*, 2021; Breuning *et al.*, 2021; Camerlink *et al.*, 2021; Candido *et al.*, 2021; Deryugina, Shurchkov and Stearns, 2021; Diaz *et al.*, 2021; Ellinas *et al.*, 2021; Gao *et al.*, 2021; Ghaffarizadeh *et al.*, 2021; Guintivano, Dick and Bulik, 2021; Hoggarth *et al.*, 2021; Krukowski, Jagsi and Cardel, 2021; Maguire *et al.*, 2021; Plaunova *et al.*, 2021; Shalaby, Allam and Buttorff, 2021; Staniscuaski *et al.*, 2021; Yildirim and Eslen-Ziya, 2021; Davis *et al.*, 2022; Stenson *et al.*, 2022), and comparisons of numbers of articles submitted or published by gender before and during the pandemic (Amano-Patiño *et al.*, 2020; Andersen *et al.*, 2020; Bell and Green, 2020; Cushman, 2020; Inno, Rotundi and Piccialli, 2020; Kibbe, 2020; Vincent-Lamarre, Sugimoto and Larivière, 2020; Wehner, Li and Nead, 2020; Bell and Fong, 2021; Biondi *et al.*, 2021; Cook *et al.*, 2021; DeFilippis *et al.*, 2021; Forti, Solino and Szabo, 2021; Fox and Meyer, 2021; Gayet-Ageron *et al.*, 2021; Gerding *et al.*, 2021; Ipe *et al.*, 2021; Jemielniak, Sławska and Wilamowski, 2021; Jordan and Carlezon, 2021; King and Frederickson, 2021; Lerchenmüller *et al.*, 2021; Mogensen, Lee and Carlos, 2021; Muric *et al.*, 2021; Nguyen *et al.*, 2021; Quak *et al.*, 2021; Ribarovska *et al.*, 2021; Squazzoni *et al.*, 2021; Williams II *et al.*, 2021; Anabaraonye *et al.*, 2022; Ayyala and Trout, 2022; Chen and Seto, 2022; Cui, Ding and Zhu, 2022; Harris *et al.*, 2022; Wooden and Hanson, 2022).

### Extracting variables

**Effect size**: We extracted values needed to calculate 130 effect sizes from 55 texts investigating the effect of the pandemic on academic research productivity between genders, before and during the pandemic, using time periods chosen according to authors’ discretion. We calculated our own effect sizes wherever possible using the available summary statistics and/or statistical inferences. For 10 effect sizes which did not have data available to calculate our own effect sizes, we used already calculated percentage changes in the gender gap in academic productivity as predicted from lasso regression (N = 2), Somers’ delta (N = 2), ordered logistic regression (N = 1) and mixed-effect models (N = 5). For 120 effect sizes, we entered summary data (N=117) or statistical inferences (N=3) into Campbell collaboration’s effect size calculator (Wilson, 2019) to calculate a standardised mean difference (d) effect size. For effect sizes calculated using summary data, 98 relied on the proportion of raw numbers of female and male authors before and after the pandemic, and 18 on the mean changes and standard deviations or standard errors in research productivity changes during the pandemic for female and male researchers. For effect sizes calculated from reported statistical tests, one converted the f-test statistic and sample size from a general linear model investigating the effect of gender on perceived work production, one converted the chi-square comparing proportions of female and male academics that experienced productivity changes due to the pandemic, and one converted the p-values from a t-test comparing mean changes in research time due to the pandemic. 10 effect sizes were calculated by obtaining raw numbers from graphs, estimated using Adobe Acrobat’s measure tool (see Supplementary file 2). Two effect sizes (Jemielniak, Sławska and Wilamowski, 2021; Stenson *et al.*, 2022) were calculated using sample sizes obtained by personal correspondence. We calculated multiple effect sizes from one study if they were for different research fields or authorship positions. We set the sign for effect sizes as negative if the pandemic had reduced relative research productivity of women (increased gender gap) and positive if the pandemic had increased the relative research productivity of women (reduced gender gap). A subset of 59 effect sizes were double-checked by A.C., A.M. and D.L and inconsistencies were discussed to ensure repeatability. K.L. then extracted the remaining 71 effect sizes.

**Variance**: Of 10 effect sizes already calculated in the original studies, 7 provided variance as the standard error, which we squared to obtain the variance; and 3 provided the variance as 95% confidence intervals, which we divided by 1.96 and then squared (Nakagawa *et al.*, 2022). For the other 120 effect sizes, variance was estimated in the Campbell collaboration calculator (Wilson, 2019) when calculating effect sizes.

**Research productivity measure**: We first recorded whether the change in research productivity was measured from survey responses (survey response studies, N=23 effect sizes) or from the number of articles submitted or published (article-output studies, N=107). Survey studies measured change in research productivity during the pandemic for each gender based on academics self-reporting their gender and change in general productivity (N=11), number of submissions (n=5), research time (N=4), number of projects (N=1), burn-out (N=1), or job loss (N=1). As 5 survey-studies measured research productivity in the number of submissions, we included these studies in the articles submitted and published category. This resulted in 18 effect sizes from surveys measuring some aspect of research productivity, 64 effect sizes measuring numbers of article submissions, and 48 effect sizes measuring numbers of publications.

**Research field**: For the article-output studies (N=107), we recorded the research field sampled based on the description in the original studies as either Medicine (N=44), Technology, Engineering, Mathematics, Chemistry and Physics (N=20), Social sciences (N=16), Biological sciences (N=17), or Multidisciplinary (N=10), following the classification scheme of (Astegiano, Sebastián-González and Castanho, 2019).

**Previous gender disparity**: For the article-output studies with available data (N=99), we recorded the number of female and male authors before the pandemic, as defined by the time-period sampled in the respective study and use this ratio as a proxy for size of the previous gender disparity in that population sampled.

**Authorship position**: For the article-output studies (N=107), we recorded whether first (N=54), middle (N=3), last (N=21), corresponding (N=15), or the total number of (N=14) authors were studied. We classified one effect size studying submitting authors, as studying corresponding authors (Fox *et al.*, 2016) and two effect sizes studying sole authors as studying last authors (Moore and Griffin, 2006).

We also extracted data for the following variables but did not include them in our analysis: timeframe before the pandemic; timeframe during the pandemic; geographic region; data availability for gender and geographic region interaction effect; data availability for gender and career stage/age interaction effect; data availability for gender and parent status interaction effect; gender assignment accuracy threshold for article-output studies and gender inference method used (Supplementary file 2, see also Limitations). We extracted these data, to offer scope for future potential analyses, but did not compare these moderators because we either had no a priori hypotheses to investigate these or were limited by sample sizes.

### Analyses

We conducted all analyses in R 3.6.2 (R Core Team, 2022). We used the ‘metafor’ package 3.0.2 to fit models, and build funnel plots (Viechtbauer, 2010). We used ‘orchaRd’ 0.0.0.9000 to build orchard plots to visualise distribution of effect sizes (points) and their precision (point size), calculated as a function of standard error (Nakagawa *et al.*, 2021).

We fitted separate models for each prediction. All models included the identity of the article the effect size was extracted from as a random effect to control for dependency in effect sizes obtained from the same study. Models that include moderators use an omnibus test of parameters reported as a QM metric, which tests the null hypothesis that all moderator effect sizes are equal and is significant when at least two moderators are different. We tested prediction 1a in a model investigating the overall effect size and we displayed this as an orchard plot. We then tested prediction 1b in a model investigating the method of measuring research productivity (survey responses, number of submissions and number of publications) as a moderator of effect size and displayed this as an orchard plot. We included the outlier (Jemielniak, Sławska and Wilamowski, 2021) in the funnel plot of article-output studies because this effect size was obtained by personal correspondence clarifying the sample sizes used in the study, which we assume was verified. We tested prediction 2a in a model investigating research field as a moderator of effect size for article-output studies in a model and displayed this as an orchard plot. We tested prediction 2b in a model investigating how previous gender disparity in research productivity before the pandemic, as measured by the proportion of female authors, influenced effect size and displayed this as a line graph, grouped by research field. To test prediction 3a, we tested in a model authorship position as a moderator on effect size for publication studies. We tested for publication bias by performing a multilevel regression model (Nakagawa *et al.*, 2022) which correlates standard error with effect size, including research productivity measure as a moderator and display this relationship in funnel plots. We tested for total heterogeneity (*I2*) using the ‘i2\_ml’ function in ‘orchaRd’. We applied a sensitivity analysis testing prediction 1a (overall pandemic effect on gender gap) and prediction 1b (method of research productivity effect on pandemic gender gap) excluding seven effect sizes using four measures of productivity from survey-based studies that are less directly comparable: research time (N = 4), job-loss (N = 1), burnout (N = 1) and the number of projects (N = 1). We also performed a leave-one-out analysis using the ‘leave1out’ function in ‘metafor’, which repeatedly fitted the model in prediction 1a, leaving out one effect size at a time.

### Limitations

Our focus is on comparing the effect of the pandemic on women relative to men. We recognize that gender extends beyond this comparison, and that biases are even more likely to target individuals whose identities are less represented and often ignored. These biases present in a lack of study of the full diversity of gender precluded our ability to include these effects in a meta-analysis. While several of the surveys we include had the option for respondents to identify beyond the binary women/men, none of these studies report on these individuals, presumably because of the respective small samples. In addition, all studies using numbers of submissions or publications (38 out of 55) to measure research productivity used automatic approaches that are more likely to mis-gender individuals as they inferred binary gender based on first names. While these approaches seemingly offer the potential to identify trends in larger samples, they themselves introduce and reinforce biases in relation to gender that are hard to assess, intersecting with biases in ethnicity as these approaches are often restricted to names common in English speaking countries (Mihaljević *et al.*, 2019). We do not include grey literature databases in our searches, which may bias our samples to those with positive effects. Most studies explored academic populations Worldwide (N = 99), or from Western (N = 28) regions, but not the Global South (N = 3) limiting investigation of interaction effects between geographic regions. Although 22/130 effect sizes from 8/55 studies held data subdivided between geographic regions, we did not extract separate effect sizes as they differed in the scale of geographic region sampled, which limited our ability to make geographic comparisons. Conclusions from survey studies are also limited to North American and Western European, since 18/23 studies are exclusive to or have the majority of respondents from these regions. We recognise there are differences between article-output studies in the length of time considered as before the pandemic (mean = 10.8 months, standard deviation = 9.9 months, range = 1-50 months) and during the pandemic (mean = 6.6 months, standard deviation = 4.6 months, range = 1-17 months). Survey studies were fielded at different times, (mean = 21/08/2020, standard deviation = 98.8 days, range = 20/04/2020 – 28/02/21) which potentially effects participants’ beliefs of productivity changes. Investigating research field and authorship position effects is limited by the unequal and sometimes small sample sizes of variables that are compared. We use 10 studies that provide effect sizes without raw data available to calculate our own effect size and recognise their differing modelling techniques may contribute to the estimated effect sizes. The patterns we describe should be seen as a potential indication that biases exist, but alternative approaches are needed to speculate about potential underlying causes and remedies.

# Acknowledgements

We express gratitude for papers not behind paywalls.

# Additional information and declarations

### Registration and protocol

This study was not registered.

### Funding

The authors received no funding for this work.

### Competing interests

The authors declare there are no competing interests.

### Author contributions

Adele Mennerat (AM) was involved in conceptualization, data extraction, methodology, analysis and reviewing and editing the manuscript draft.

Antica Culina (AC) was involved in conceptualization, data extraction, methodology, analysis and reviewing and editing the manuscript draft.

Dieter Lukas (DL) was involved in conceptualization, data extraction, methodology, analysis and reviewing and editing the manuscript draft.

Hannah Dugdale (HD) was involved in conceptualization, methodology, analysis and reviewing and editing the manuscript draft.

Kiran Lee (KL) was involved in conceptualization, data extraction, methodology, analysis, writing the first draft and reviewing and editing the draft.

### Data availability

Data and materials to reproduce the meta-analysis can be found at Zenodo [DOI: 10.5281/zenodo.7517177](https://doi.org/10.5281/zenodo.7517177).

### Supplementary information

Supplemental information is currently at the end of this document but can be separated if requested.

# References

Abdellatif, A. and Gatto, M. (2020) ‘It’s OK not to be OK: Shared reflections from two PhD parents in a time of pandemic’, *Gender, Work & Organization*, 27(5), pp. 723–733. Available at: https://doi.org/10.1111/gwao.12465.

Altan-Olcay, Ö. and Bergeron, S. (2022) ‘Care in times of the pandemic: Rethinking work meanings of work in the university’, *Gender, Work & Organization* [Preprint]. Available at: https://doi.org/10.1111/gwao.12871.

Amano-Patiño, N. *et al.* (2020) ‘The Unequal Effects of Covid-19 on Economists’ Research Productivity’. Available at: https://doi.org/10.17863/CAM.57979.

Anabaraonye, N. *et al.* (2022) ‘Impact of the early COVID-19 pandemic on gender participation in academic publishing in radiation oncology’, *Advances in Radiation Oncology*, 7(2), p. 100845. Available at: https://doi.org/10.1016/j.adro.2021.100845.

Andersen, J.P. *et al.* (2020) ‘COVID-19 medical papers have fewer women first authors than expected’, *elife*, 9. Available at: https://doi.org/10.7554/eLife.58807.

Anwer, M. (2020) ‘Academic labor and the global pandemic: Revisiting life-work balance under COVID-19’, *Susan Bulkeley Butler Center for leadership excellence and advance working paper series*, 3(1), pp. 5–13. Available at: https://www.purdue.edu/butler/working-paper-series/2020/special%20new.php.

Astegiano, J., Sebastián-González, E. and Castanho, C.D.T. (2019) ‘Unravelling the gender productivity gap in science: a meta-analytical review’, *Royal Society Open Science*, 6(6). Available at: https://doi.org/10.1098/RSOS.181566.

Ayyala, R.S. and Trout, A.T. (2022) ‘Gender trends in authorship of Pediatric Radiology publications and impact of the COVID-19 pandemic’, *Pediatric Radiology*, 52(5), pp. 868–873. Available at: https://doi.org/10.1007/s00247-021-05213-6.

Babcock, L. *et al.* (2017) ‘Gender differences in accepting and receiving requests for tasks with low promotability’, *American Economic Review*, 107(3), pp. 714–47. Available at: https://doi.org/10.1257/aer.20141734.

Babcock, L. *et al.* (2022) ‘Saying “no” in science isn’t enough’, *Nature* [Preprint]. Available at: https://doi.org/10.1038/d41586-022-03677-6.

Barber, B.M. *et al.* (2021) ‘What explains differences in finance research productivity during the pandemic?’, *The Journal of Finance*, 76(4), pp. 1655–1697. Available at: https://doi.org/10.1111/jofi.13028.

Bell, K. and Green, J. (2020) ‘Premature evaluation? Some cautionary thoughts on global pandemics and scholarly publishing.’, *Critical Public Health*, 30(4), pp. 379–383. Available at: https://doi.org/10.1080/09581596.2020.1769406.

Bell, M.L. and Fong, K.C. (2021) ‘Gender differences in first and corresponding authorship in public health research submissions during the COVID-19 pandemic’, *American Journal of Public Health*, 111(1), pp. 159–163. Available at: https://doi.org/10.2105/AJPH.2020.305975.

van den Besselaar, P. and Sandström, U. (2016) ‘Gender differences in research performance and its impact on careers: a longitudinal case study’, *Scientometrics*, 106(1), pp. 143–162. Available at: https://doi.org/10.1007/s11192-015-1775-3.

van den Besselaar, P. and Sandström, U. (2017) ‘Vicious circles of gender bias, lower positions, and lower performance: Gender differences in scholarly productivity and impact’, *PLOS ONE*, 12(8), p. e0183301. Available at: https://doi.org/10.1371/journal.pone.0183301.

Biondi, B. *et al.* (2021) ‘Journal submissions, review and editorial decision patterns during initial COVID-19 restrictions’, *Food Policy*, 105, p. 102167. Available at: https://doi.org/10.1016/j.foodpol.2021.102167.

Boncori, I. (2020) ‘The Never-ending Shift: A feminist reflection on living and organizing academic lives during the coronavirus pandemic’, *Gender, Work & Organization*, 27(5), pp. 677–682. Available at: https://doi.org/10.1111/gwao.12451.

Breuning, M. *et al.* (2021) ‘The great equalizer? Gender, parenting, and scholarly productivity during the global pandemic’, *PS: Political Science & Politics*, 54(3), pp. 427–431. Available at: https://doi.org/10.1017/S1049096520002036.

Camerlink, I. *et al.* (2021) ‘Impacts of the COVID-19 pandemic on animal behaviour and welfare researchers.’, *Applied Animal Behaviour Science*, 236, p. 105255. Available at: https://doi.org/10.1016/j.applanim.2021.105255.

Candido, M.R. *et al.* (2021) ‘Social sciences in the COVID-19 pandemic: work rotuines and inequalities’, *Sociologia & Antropologia*, 11, pp. 31–65. Available at: https://doi.org/10.1590/2238-38752021v11esp2.

Carli, L.L. (2020) ‘Women, Gender equality and COVID-19’, *Gender in Management: An International Journal* [Preprint]. Available at: https://doi.org/10.1108/GM-07-2020-0236.

Carlson, D.L., Petts, R.J. and Pepin, J.R. (2020) ‘US Couples’ Divisions of Housework and Childcare during COVID-19 Pandemic’. Available at: https://doi.org/10.31235/osf.io/jy8fn.

Chen, T.-H.K. and Seto, K.C. (2022) ‘Gender and authorship patterns in urban land science’, *Journal of Land Use Science*, 17(1), pp. 245–261. Available at: https://doi.org/10.1080/1747423X.2021.2018515.

Clark, D. (2020) ‘Reflections on institutional equity for faculty in response to COVID-19’, *Susan Bulkeley Butler Center for Leadership Excellence and ADVANCE Working Paper Series*, 3(2). Available at: https://par.nsf.gov/servlets/purl/10280774 (Accessed: 20 September 2021).

Cook, J. *et al.* (2021) ‘Gender differences in authorship of obstetrics and gynecology publications during the coronavirus disease 2019 pandemic’, *American journal of obstetrics & gynecology MFM*, 3(1). Available at: https://doi.org/10.1016/j.ajogmf.2020.100268.

Corbera, E. *et al.* (2020) ‘Academia in the Time of COVID-19: Towards an Ethics of Care’, *Planning Theory and Practice*, 21(2), pp. 191–199. Available at: https://doi.org/10.1080/14649357.2020.1757891.

Cui, R., Ding, H. and Zhu, F. (2022) ‘Gender inequality in research productivity during the COVID-19 pandemic’, *Manufacturing & Service Operations Management*, 24(2), pp. 707–726. Available at: https://doi.org/10.1287/msom.2021.0991.

Cushman, M. (2020) ‘Gender gap in women authors is not worse during COVID-19 pandemic: Results from Research and Practice in Thrombosis and Haemostasis’, *Research and Practice in Thrombosis and Haemostasis*, 4(5), p. 672. Available at: https://doi.org/10.1002/rth2.12399.

Davis, J.C. *et al.* (2022) ‘Are we failing female and racialized academics? A Canadian national survey examining the impacts of the COVID-19 pandemic on tenure and tenure-track faculty’, *Gender, Work & Organization*, 29(3), pp. 703–722. Available at: https://doi.org/10.1111/gwao.12811.

Day, A.E., Corbett, P. and Boyle, J. (2020) ‘Is there a gender gap in chemical sciences scholarly communication?’, *Chemical Science*, 11(8), pp. 2277–2301. Available at: https://doi.org/10.1039/c9sc04090k.

DeFilippis, E.M. *et al.* (2021) ‘Gender differences in publication authorship during COVID-19: a bibliometric analysis of high-impact cardiology journals’, *Journal of the American Heart Association*, 10(5), p. e019005. Available at: https://doi.org/10.1161/JAHA.120.019005.

Deryugina, T., Shurchkov, O. and Stearns, J.E. (2021) ‘COVID-19 Disruptions Disproportionately Affect Female Academics’. Available at: https://doi.org/10.1257/pandp.20211017.

Diaz, J.O. *et al.* (2021) ‘Burnout syndrome in pediatric urology: A perspective during the COVID-19 pandemic—Ibero-American survey’, *Journal of Pediatric Urology*, 17(3), pp. 402-e1. Available at: https://doi.org/10.1016/j.jpurol.2021.01.015.

Docka-Filipek, D. and Stone, L.B. (2021) ‘Twice a “housewife”: On academic precarity,“hysterical” women, faculty mental health, and service as gendered care work for the “university family” in pandemic times’, *Gender, Work & Organization*, 28(6), pp. 2158–2179. Available at: https://doi.org/10.1111/gwao.12723.

Ellinas, E.H. *et al.* (2021) ‘Winners and Losers in Academic Productivity During the COVID-19 Pandemic: Is the Gender Gap Widening for Faculty?’, *Journal of women’s health (2002)* [Preprint]. Available at: https://doi.org/10.1089/jwh.2021.0321.

Else, H. (2020) ‘How a torrent of COVID science changed research publishing–in seven charts’, *Nature*, 588(7839), pp. 553–554. Available at: https://doi.org/10.1038/d41586-020-03564-y.

Foo, Y.Z. *et al.* (2021) ‘A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution’, *Methods in Ecology and Evolution*, 12(9), pp. 1705–1720. Available at: https://doi.org/10.1111/2041-210X.13654.

Forti, L.R., Solino, L.A. and Szabo, J.K. (2021) ‘Trade-off between urgency and reduced editorial capacity affect publication speed in ecological and medical journals during 2020’, *Humanities and Social Sciences Communications*, 8(1), pp. 1–9. Available at: https://doi.org/10.1057/s41599-021-00920-9.

Fortin, J. and Taylor, D.B. (2020) ‘Florida State University Child Care Policy Draws Backlash’, *The New York Times*, 2 July. Available at: https://www.nytimes.com/2020/07/02/us/fsu-telecommute-remote.html (Accessed: 6 November 2022).

Fox, C.W. *et al.* (2016) ‘Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal’, *Functional Ecology*, 30(1), pp. 126–139. Available at: https://doi.org/10.1111/1365-2435.12587.

Fox, C.W. and Meyer, J. (2021) ‘The influence of the global COVID-19 pandemic on manuscript submissions and editor and reviewer performance at six ecology journals’, *Functional Ecology*. Wiley Online Library.

Fox, C.W. and Paine, C.T. (2019) ‘Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution’, *Ecology and Evolution*, 9(6), pp. 3599–3619. Available at: https://doi.org/10.1002/ece3.4993.

Gao, J. *et al.* (2021) ‘Potentially long-lasting effects of the pandemic on scientists’, *Nature communications*, 12(1), pp. 1–6. Available at: https://doi.org/10.1038/s41467-021-26428-z.

Gayet-Ageron, A. *et al.* (2021) ‘Female authorship of covid-19 research in manuscripts submitted to 11 biomedical journals: cross sectional study.’, *BMJ (Clinical research ed.)*, 375, p. n2288. Available at: https://doi.org/10.1136/bmj.n2288.

Gerding, A.B. *et al.* (2021) ‘Scholarly productivity in clinical pharmacology amid pandemic-related workforce disruptions: are men and women affected equally?’, *Clinical Pharmacology and Therapeutics*, 110(4), p. 841. Available at: https://doi.org/10.1002/cpt.2358.

Ghaffarizadeh, S. Aryan *et al.* (2021) ‘Life and work of researchers trapped in the COVID-19 pandemic vicious cycle’, *bioRxiv*, p. 2021.02.02.429476. Available at: https://doi.org/10.1101/2021.02.02.429476.

Grames, E.M. *et al.* (2019) ‘An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks’, *Methods in Ecology and Evolution*, 10(10), pp. 1645–1654. Available at: https://doi.org/10.1111/2041-210X.13268.

Guarino, C.M. and Borden, V.M.H. (2017) ‘Faculty Service Loads and Gender: Are Women Taking Care of the Academic Family?’, *Research in Higher Education*, 58(6), pp. 672–694. Available at: https://doi.org/10.1007/s11162-017-9454-2.

Guintivano, J., Dick, D. and Bulik, C.M. (2021) ‘Psychiatric genomics research during the COVID-19 pandemic: A survey of Psychiatric Genomics Consortium researchers’, *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 186(1), pp. 40–49. Available at: https://doi.org/10.1002/ajmg.b.32838.

Guy, B. and Arthur, B. (2020) ‘Academic motherhood during COVID-19: Navigating our dual roles as educators and mothers’, *Gender, Work & Organization*, 27(5), pp. 887–899. Available at: https://doi.org/10.1111/gwao.12493.

Hagan, A.K. *et al.* (2020) ‘Women Are Underrepresented and Receive Differential Outcomes at ASM Journals: a Six-Year Retrospective Analysis’, *mBio*, 11(6), pp. e01680-20. Available at: https://doi.org/10.1128/mBio.01680-20.

Harris, B. *et al.* (2022) ‘Exploratory Investigation of Gender Differences in School Psychology Publishing Before and During the Initial Phase of COVID-19’, *Canadian Journal of School Psychology*, 37(2), pp. 204–211. Available at: https://doi.org/10.1177/08295735221074473.

Herman, E. *et al.* (2021) ‘The impact of the pandemic on early career researchers: what we already know from the internationally published literature’, *El Profesional de la Información*, 30(2). Available at: https://doi.org/10.3145/epi.2021.mar.08.

Hoggarth, J.A. *et al.* (2021) ‘Impacts of the COVID-19 pandemic on women and early career archaeologists’, *Heritage*, 4(3), pp. 1681–1702. Available at: https://doi.org/10.3390/heritage4030093.

Horbach, S.P.J.M. (2020) ‘Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19’, *Quantitative Science Studies*, 1(3), pp. 1056–1067. Available at: https://doi.org/10.1162/QSS\_A\_00076.

Huang, J. *et al.* (2020) ‘Historical comparison of gender inequality in scientific careers across countries and disciplines’, *Proceedings of the National Academy of Sciences*, 117(9), pp. 4609–4616. Available at: https://doi.org/10.1073/PNAS.1914221117.

Inno, L., Rotundi, A. and Piccialli, A. (2020) ‘COVID-19 lockdown effects on gender inequality’, *Nature Astronomy*, 4(12), pp. 1114–1114. Available at: https://doi.org/10.1038/s41550-020-01258-z.

Ipe, T.S. *et al.* (2021) ‘The impact of COVID-19 on academic productivity by female physicians and researchers in transfusion medicine’, *Transfusion*, 61(6), pp. 1690–1693. Available at: https://doi.org/10.1111/trf.16306.

Jemielniak, D., Sławska, A. and Wilamowski, M. (2021) ‘COVID-19 effect on the gender gap in academic publishing’, *Journal of Information Science*, p. 01655515211068168. Available at: https://doi.org/10.1177/01655515211068168.

Jordan, C.J. and Carlezon, W.A. (2021) ‘Effects of the COVID-19 pandemic on gender representation among corresponding authors of Neuropsychopharmacology (NPP) manuscripts: submissions during January–June, 2020’, *Neuropsychopharmacology*, 46(2), pp. 269–270. Available at: https://doi.org/10.1038/s41386-020-00869-4.

Kaatz, A., Gutierrez, B. and Carnes, M. (2014) ‘Threats to objectivity in peer review: the case of gender’, *Trends in pharmacological sciences*, 35(8), pp. 371–373. Available at: https://doi.org/10.1016/j.tips.2014.06.005.

Kibbe, M.R. (2020) ‘Consequences of the COVID-19 Pandemic on Manuscript Submissions by Women’, *JAMA Surgery*, 155(9), pp. 803–804. Available at: https://doi.org/10.1001/jamasurg.2020.3917.

King, M.M. and Frederickson, M.E. (2021) ‘The Pandemic Penalty: The Gendered Effects of COVID-19 on Scientific Productivity’:, *Socius* [Preprint]. Available at: https://doi.org/10.1177/23780231211006977.

Krukowski, R.A., Jagsi, R. and Cardel, M.I. (2021) ‘Academic productivity differences by gender and child age in science, technology, engineering, mathematics, and medicine faculty during the COVID-19 pandemic’, *Journal of Women’s Health*, 30(3), pp. 341–347. Available at: https://doi.org/10.1089/jwh.2020.8710.

Landis, J. and Koch, G. (1977) ‘The measurement of observer agreement for categorical data.’, *Biometrics*, 33(1), pp. 159–174. Available at: https://doi.org/10.2307/2529310.

Lerchenmüller, C. *et al.* (2021) ‘Longitudinal analyses of gender differences in first authorship publications related to COVID-19’, *BMJ open*, 11(4), p. e045176. Available at: https://doi.org/10.1136/bmjopen-2020-045176.

Livoreil, B. *et al.* (2017) ‘Systematic searching for environmental evidence using multiple tools and sources’, *Environmental Evidence 2017 6:1*, 6(1), pp. 1–14. Available at: https://doi.org/10.1186/S13750-017-0099-6.

Macaluso, B. *et al.* (2016) ‘Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship’, *Academic Medicine*, 91(8), pp. 1136–1142. Available at: https://doi.org/10.1097/ACM.0000000000001261.

Madsen, E.B. *et al.* (2022) ‘Author-level data confirm the widening gender gap in publishing rates during COVID-19’, *eLife*. Edited by P. Rodgers, 11, p. e76559. Available at: https://doi.org/10.7554/eLife.76559.

Maguire, R. *et al.* (2021) ‘Research interrupted: The impact of the COVID-19 pandemic on multiple sclerosis research in the field of rehabilitation and quality of life’, *Multiple Sclerosis Journal–Experimental, Translational and Clinical*, 7(3), p. 20552173211038030. Available at: https://doi.org/10.1177/20552173211038030.

McHugh, M.L. (2012) ‘Interrater reliability: the kappa statistic’, *Biochemia Medica*, 22(3), p. 276. Available at: https://doi.org/10.11613/BM.2012.031.

Mihaljević, H. *et al.* (2019) ‘Reflections on Gender Analyses of Bibliographic Corpora’, *Frontiers in Big Data*, 2. Available at: https://doi.org/10.3389/fdata.2019.00029.

Minello, A., Martucci, S. and Manzo, L.K. (2021) ‘The pandemic and the academic mothers: present hardships and future perspectives’, *European Societies*, 23(sup1), pp. S82–S94. Available at: https://doi.org/10.1080/14616696.2020.1809690.

Mitchell, S.M. and Hesli, V.L. (2013) ‘Women don’t ask? Women don’t say no? Bargaining and service in the political science profession’, *PS: Political Science & Politics*, 46(2), pp. 355–369. Available at: https://doi.org/10.1017/S1049096513000073.

Mogensen, M.A., Lee, C.I. and Carlos, R.C. (2021) ‘The Impact of the COVID-19 Pandemic on Journal Scholarly Activity Among Female Contributors’, *Journal of the American College of Radiology*, 18(7), pp. 1044–1047. Available at: https://doi.org/10.1016/j.jacr.2021.01.011.

Moher, D. *et al.* (2009) ‘Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement’, *PLOS Medicine*, 6(7), p. e1000097. Available at: https://doi.org/10.1371/JOURNAL.PMED.1000097.

Moore, M.T. and Griffin, B.W. (2006) ‘Identification of factors that influence authorship name placement and decisions to collaborate in peer-reviewed, education-related publications’, *Studies in Educational Evaluation*, 32(2), pp. 125–135. Available at: https://doi.org/10.1016/j.stueduc.2006.04.004.

Muric, G. *et al.* (2021) ‘Gender disparity in the authorship of biomedical research publications during the COVID-19 pandemic: retrospective observational study’, *Journal of medical Internet research*, 23(4), p. e25379. Available at: https://doi.org/10.2196/25379.

Murray, D. *et al.* (2019) ‘Author-reviewer homophily in peer review’, *BioRxiv*, p. 400515. Available at: https://doi.org/10.1101/400515.

Myers, K.R. *et al.* (2020) ‘Unequal effects of the COVID-19 pandemic on scientists’, *Nature human behaviour*, 4(9), pp. 880–883. Available at: https://doi.org/10.1038/s41562-020-0921-y.

Nakagawa, S. *et al.* (2021) ‘The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond’, in *Research Synthesis Methods*. John Wiley and Sons Ltd, pp. 4–12. Available at: https://doi.org/10.1002/jrsm.1424.

Nakagawa, S. *et al.* (2022) ‘Methods for testing publication bias in ecological and evolutionary meta-analyses’, *Methods in Ecology and Evolution*, 13(1), pp. 4–21. Available at: https://doi.org/10.1111/2041-210X.13724.

Nash, M. and Churchill, B. (2020) ‘Caring during COVID-19: A gendered analysis of Australian university responses to managing remote working and caring responsibilities’, *Gender, Work & Organization*, 27(5), pp. 833–846. Available at: https://doi.org/10.1111/gwao.12484.

National Academies of Sciences, E. (2021) *The Impact of COVID-19 on the Careers of Women in Academic Sciences, Engineering, and Medicine*. Available at: https://doi.org/10.17226/26061.

Nguyen, A.X. *et al.* (2021) ‘Impact of COVID-19 on longitudinal ophthalmology authorship gender trends’, *Graefe’s Archive for Clinical and Experimental Ophthalmology*, 259(3), pp. 733–744. Available at: https://doi.org/10.1007/s00417-021-05085-4.

O’Meara, K., Kuvaeva, A. and Nyunt, G. (2017) ‘Constrained choices: A view of campus service inequality from annual faculty reports’, *The Journal of Higher Education*, 88(5), pp. 672–700. Available at: https://doi.org/10.1080/00221546.2016.1257312.

O’Meara, K.A. *et al.* (2017) ‘Asked More Often: Gender Differences in Faculty Workload in Research Universities and the Work Interactions That Shape Them’, *American Educational Research Journal*, 54(6), pp. 1154–1186. Available at: https://doi.org/10.3102/0002831217716767.

Ouzzani, M. *et al.* (2016) ‘Rayyan—a web and mobile app for systematic reviews’, *Systematic Reviews 2016 5:1*, 5(1), pp. 1–10. Available at: https://doi.org/10.1186/S13643-016-0384-4.

Plaunova, A. *et al.* (2021) ‘Impact of COVID-19 on radiology faculty-an exacerbation of gender differences in unpaid home duties and professional productivity’, *Academic Radiology*, 28(9), pp. 1185–1190. Available at: https://doi.org/10.1016/j.acra.2021.05.004.

Porter, S.R. (2007) ‘A closer look at faculty service: What affects participation on committees?’, *The Journal of Higher Education*, 78(5), pp. 523–541. Available at: https://doi.org/10.1353/jhe.2007.0027.

Powell, K. (2016) ‘Does it take too long to publish research?’, *Nature*, 530(7589), pp. 148–151. Available at: https://doi.org/10.1038/530148a.

Quak, E. *et al.* (2021) ‘Author Gender Inequality in Medical Imaging Journals and the COVID-19 Pandemic’, *Radiology*, 300(1), pp. E301–E307. Available at: https://doi.org/10.1148/radiol.2021204417.

R Core Team (2022) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Reuben, E., Sapienza, P. and Zingales, L. (2014) ‘How stereotypes impair women’s careers in science’, *Proceedings of the National Academy of Sciences*, 111(12), pp. 4403–4408. Available at: https://doi.org/10.1073/pnas.1314788111.

Ribarovska, A.K. *et al.* (2021) ‘Gender inequality in publishing during the COVID-19 pandemic’, *Brain, behavior, and immunity*, 91, p. 1. Available at: https://doi.org/10.1016/j.bbi.2020.11.022.

Rodríguez-Rivero, R. *et al.* (2020) ‘Is it time for a revolution in work–life balance? Reflections from Spain’, *Sustainability*, 12(22), p. 9563. Available at: https://doi.org/10.3390/su12229563.

Schiebinger, L. and Gilmartin, S.K. (2010) ‘Housework is an academic issue’, *Academe*, 96(1), pp. 39–44. Available at: https://www.aaup.org/article/housework-academic-issue#.Y53t1ezP3MJ.

Schiebinger, L.L., Henderson, A.D. and Gilmartin, S.K. (2008) *Dual-career academic couples: What universities need to know*. Michelle R. Clayman institute for gender research, Stanford University. Available at: https://gender.stanford.edu/sites/gender/files/dualcareerfinal\_0.pdf (Accessed: 27 August 2022).

Shalaby, M., Allam, N. and Buttorff, G.J. (2021) ‘Leveling the field: Gender inequity in academia during COVID-19’, *PS: Political Science & Politics*, 54(4), pp. 661–667. Available at: https://doi.org/10.1017/S1049096521000615.

Squazzoni, F. *et al.* (2021) ‘Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals’, *PloS one*, 16(10), p. e0257919. Available at: https://doi.org/10.1371/journal.pone.0257919.

Staniscuaski, F. *et al.* (2021) ‘Gender, race and parenthood impact academic productivity during the COVID-19 pandemic: from survey to action’, *Frontiers in psychology*, 12, p. 663252. Available at: https://doi.org/10.3389/fpsyg.2021.663252.

Stenson, M.C. *et al.* (2022) ‘Impact of COVID-19 on access to laboratories and human participants: exercise science faculty perspectives’, *Advances in Physiology Education*, 46(2), pp. 211–218. Available at: https://doi.org/10.1152/advan.00146.2021.

Viechtbauer, W. (2010) ‘Conducting Meta-Analyses in R with the metafor Package’, *Journal of Statistical Software*, 36(3), pp. 1–48. Available at: https://doi.org/10.18637/JSS.V036.I03.

Vincent-Lamarre, P., Sugimoto, C.R. and Larivière, V. (2020) ‘The decline of women’s research production during the coronavirus pandemic’, *Nature Index*, May 19, pp. 1–11.

Wehner, M.R., Li, Y. and Nead, K.T. (2020) ‘Comparison of the proportions of female and male corresponding authors in preprint research repositories before and during the COVID-19 pandemic’, *JAMA network open*, 3(9), pp. e2020335–e2020335. Available at: https://doi.org/10.1001/jamanetworkopen.2020.20335.

Williams II, W.A. *et al.* (2021) ‘Impact of the coronavirus disease 2019 pandemic on authorship gender in the Journal of Pediatrics: disproportionate productivity by international male researchers’, *The Journal of Pediatrics*, 231, pp. 50–54. Available at: https://doi.org/10.1016/j.jpeds.2020.12.032.

Wilson, D.B. (2019) *Practical Meta-Analysis Effect Size Calculator [Online calculator]*. Available at: https://campbellcollaboration.org/research-resources/effect-size-calculator.html (Accessed: 28 August 2022).

Wooden, P. and Hanson, B. (2022) ‘Effects of the COVID-19 Pandemic on Authors and Reviewers of American Geophysical Union Journals’, *Earth and Space Science*, 9(2), p. e2021EA002050. Available at: https://doi.org/10.1029/2021EA002050.

Yildirim, T.M. and Eslen-Ziya, H. (2021) ‘The differential impact of COVID-19 on the work conditions of women and men academics during the lockdown’, *Gender, Work & Organization*, 28, pp. 243–249. Available at: https://doi.org/10.1111/gwao.12529.

# Supplementary materials

**Table S1. Naive Boolean search string used in initial literature search.** Wildcards (\*) were used to return results containing different word endings. Texts were limited to those since 2020. Search was conducted on 10/07/2021.

|  |  |
| --- | --- |
| Concept group (PICO) | Terms |
| Academia (population) | ( academi\* OR author\* OR database\* OR journal\* OR research OR scien\* ) |
| Gender (population) | AND    ( female\* OR gender OR male\* OR men OR women ) |
| Pandemic (intervention) | AND    ( coronavirus OR covid OR pandemic ) |
| Inequality (comparator) | AND    ( bias\* OR disparit\* OR disproportion\* OR fewer OR gap OR "gender difference\*" OR imbalance\* OR inequalit\* OR inequit\* OR parity OR "sex difference\*" OR skew\* OR unequal ) |
| Productivity (outcome) | AND  ( performan\* OR publication\* OR publish\* OR productiv\* ) |
| Exclusion of biomedical studies (population) | AND NOT  ( surviv\* OR experiment OR laboratory ) |

**Table S2. Top 60 strongest terms in the term co-occurrence matrix and their rank in ascending order.** Terms in bold were included in the AND NOT operator concept group excluding biomedical studies.

|  |  |  |
| --- | --- | --- |
| Term | Strength | Rank ascending |
| child | 4257 | 2063 |
| regression | 4335 | 2064 |
| survey | 4367 | 2065 |
| ***acute respiratory*** | 4427 | 2066 |
| embase | 4637 | 2067 |
| databases | 4646 | 2068 |
| death | 4664 | 2069 |
| knowledge | 4764 | 2070 |
| hip | 4918 | 2071 |
| mortality | 5000 | 2072 |
| female | 5065 | 2073 |
| rights reserved | 5257 | 2074 |
| city | 5504 | 2075 |
| information | 5700 | 2076 |
| symptom | 6108 | 2077 |
| sars-cov-2 | 6112 | 2078 |
| sars-cov | 6185 | 2079 |
| coronavirus disease 2019 | 6397 | 2080 |
| systematic review | 6506 | 2081 |
| rest | 6578 | 2082 |
| science | 6659 | 2083 |
| distance | 6679 | 2084 |
| literature | 6816 | 2085 |
| database | 6916 | 2086 |
| medical | 7023 | 2087 |
| ***icu*** | 7170 | 2088 |
| affect | 7297 | 2089 |
| control | 7386 | 2090 |
| population | 7418 | 2091 |
| women | 7422 | 2092 |
| bias | 7665 | 2093 |
| gender | 7693 | 2094 |
| male | 7977 | 2095 |
| coronavirus disease | 8009 | 2096 |
| article | 8135 | 2097 |
| virus disease | 8230 | 2098 |
| ***iga*** | 8235 | 2099 |
| hospital | 8318 | 2100 |
| work | 8319 | 2101 |
| covid-19 pandemic | 8377 | 2102 |
| ***gis*** | 8489 | 2103 |
| severe | 8518 | 2104 |
| time | 8831 | 2105 |
| infection | 8914 | 2106 |
| ***rna*** | 10303 | 2107 |
| outcome | 10461 | 2108 |
| review | 11869 | 2109 |
| coronavirus | 12243 | 2110 |
| stem | 12316 | 2111 |
| ***risk*** | 12776 | 2112 |
| analysis | 13610 | 2113 |
| research | 13737 | 2114 |
| virus | 13835 | 2115 |
| pandemic | 14212 | 2116 |
| author | 14333 | 2117 |
| car | 15893 | 2118 |
| over | 15964 | 2119 |
| health | 16176 | 2120 |
| covid-19 | 21088 | 2121 |
| age | 21194 | 2122 |
| covid | 21330 | 2123 |

**Table S3. Flowchart of questions used for initial screen of title, abstract and key-words**. Any 'No' answer meant that the study was excluded. 'Yes' AND/OR 'Maybe' answers to all of the below meant that the study was included. Reports such as reviews or comments investigating our PICO framework were included. Studies that lacked a formal abstract but had a title or informal abstract suggesting at least 3 of the above questions were answered with a ‘yes’.

|  |  |
| --- | --- |
| Initial screen question | Additional notes |
| 1) Population: Is there suggestion the study investigates academia/ research/academics/journals? | Include studies that do not specifically investigate academia/research/academics/journals, but considers the effect on ‘work’ or ‘work productivity’ at a broader level, with potentially enough power to subdivide effects on different workplaces, such as academia.  Some academic fields such as those within medicine and engineering are dominated by a practical/applied/professional/labour component. Exclude studies with explicit reference to only investigating the professional/practical/labour context of these fields. Include studies if no reference is made to whether the academic field is investigated in a research or professional capacity. |
| 2) Intervention: Is there suggestion the study investigates the effect of the pandemic? | There is no exact start date to the global pandemic. Include all studies mentioning the pandemic as the exact dates defining the duration of the pandemic will be chosen at the discretion of the authors. |
| 3) Comparison: Is there suggestion the study investigates pre-pandemic? | Often titles and abstracts do not have dates and will not refer explicitly to pre-pandemic data being included. Include studies if there is suggestion the ‘effect of the pandemic’ (or similar) is considered, assume a comparison with pre-pandemic data is included even if no explicit reference to pre-pandemic productivity is considered.  There is no exact start date to the global pandemic. Include all studies mentioning the pandemic as the exact dates defining the duration of time before the pandemic will be chosen at the discretion of the authors. |
| 4) Outcome: Is there suggestion the study investigates the effect on gender in productivity/publication/submission/authorship numbers? | Include studies if there is no reference to investigating the effect of gender or sex, but does refer to investigating broader sociodemographic or socioeconomic factors because gender may be investigated in the full text.    Include studies that suggest investigation of any form of research-related activity. |

**Figure S1. PRISMA flow diagram for number of records included at each stage in first and iterated search.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Identification** |  |  | 10/07/2021 search | |  |  |
|  | Web of Science | Scopus | EBSCO | ProQuest |  |
|  | N=199 | N=126 | N=276 | N=99 |  |
|  |  | ↓ | |  |  |
|  |  | Total | |  |  |
|  |  | N=700 | |  |  |
| **Screening** |  |  | ↓ | |  |  |
|  |  | De-duplicated | |  |  |
|  |  | N=580 | |  |  |
|  |  | ↓ | |  |  |
|  |  | Title, abstract, keyword screened | |  |  |
|  |  | N=81 | |  |  |
| **Include** |  |  | ↓ | |  |  |
|  |  | Full text screened | |  |  |
|  |  | N=25 | |  |  |
|  |  |  |  | |  |  |
|  |  |  | 28/02/2022 search | |  |  |
| **Identification** |  | Web of Science | Scopus | EBSCO | ProQuest |  |
|  | N=413 | N=258 | N=542 | N=433 |  |
|  |  | ↓ | |  |  |
|  |  | Total | |  |  |
|  |  | N=1646 | |  |  |
|  |  | ↓ | |  |  |
| **Screening­­­­** |  |  | De-duplicated | |  |  |
|  |  | N=1208 | | → | Excluded initial screen |
|  |  | ↓ | |  | N=1039 |
|  |  | Title, abstract, keyword screened | |  |
|  |  | N=169 | | → | Excluded full screen |
|  |  | ↓ | |  | Missed 1 qualifier N=53 |
| **Included** | Scoped texts |  | Full text screened | |  | Missed 2 qualifiers N=38 |
| N=5 | → | N=50 | |  | Missed 3 qualifiers N=27 |
|  |  | ↓ | |  | Missed 4 qualifiers N=1 |
|  |  | Total texts in Meta-analysis | |  |  |
|  |  | N=55 | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |
|  |  |  |  | |  |  |

**Table S4. PRISMA checklist**

| **Section and Topic** | **Item #** | **Checklist item** | **Where item is reported** |
| --- | --- | --- | --- |
| **TITLE** | | |  |
| Title | 1 | Identify the report as a systematic review. | Page 2 |
| **ABSTRACT** | | |  |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Page 2 |
| **INTRODUCTION** | | |  |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Page 3 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 4 |
| **METHODS** | | |  |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Pages 20,21 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Pages 20,21 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Pages 20,21 & Table 1 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Pages 21 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 21 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Pages 20,21 |
| 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Pages 20,21 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Page 21 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Pages 20,21 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 21 |
| 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Page 23 |
| 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Page 24 |
| 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 24 |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Page 25 |
| 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Page 24 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Page 24 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Page 23 |
| **RESULTS** | | |  |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Table S4 |
| 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Table S4 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Figure 1 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Figure 1 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Pages 5, 7 & Figures 1-5 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Page 5 |
| 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Pages 5, 7 |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Figure 2 |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Figures 2, 3, 5 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Page 15 & Figure 6 |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Figures 1-6 |
| **DISCUSSION** | | |  |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Pages 17, 18 |
| 23b | Discuss any limitations of the evidence included in the review. | Page 17 |
| 23c | Discuss any limitations of the review processes used. | Pages 18, 25 |
| 23d | Discuss implications of the results for practice, policy, and future research. | Page 19 |
| **OTHER INFORMATION** | | |  |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | Page 26 |
| 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | NA |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | NA |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | Page 26 |
| Competing interests | 26 | Declare any competing interests of review authors. | Page 26 |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | Page 27 |